Challenges in the Inter-disciplinary Use of Comparative Law

Christoph Engel

MAX PLANCK SOCIETY

Challenges in the Interdisciplinary Use of Comparative Law

Christoph Engel

November 2020

I. Introduction

Interdisciplinarity is akin to foreign trade. If a country with a well-educated and equally moti-vated labour force imports food, this does not mean that there could no food be grown in that country. Its population did not starve when international trade was less developed. But the make-up of its labour force gives the country a competitive advantage. If the labour force of the food producing country is less advanced, both stand to gain from the exchange. The former country may devote precious human capital to purposes that are more productive than grow-ing plants. And the latter country gets access to goods or services that require sophisticated input. This is the core of the theory of comparative advantage.[1]

Arguably, comparative law and the social sciences are in the same situation. If a social scien-tist is interested in cross-country comparisons, she could of course start the enterprise by building her own dataset. As most of social life is determined, shaped, or at least moderated by the law, the quality of such data would benefit from taking legal factors into account. As law to a large extent is still a national affair, and hardly any rule is uniform across the world, there is interesting variance which the social scientist may want to exploit. Yet few social sci-entists have legal training. There is a palpable risk that they will misinterpret the legal factors of interest. Of course, individual social scientists could specialise on these factors, and be-come better over time. But there is a rich field of comparative law that has already built this expertise. Belabouring the metaphor, this expertise seems to be the comparative advantage of legal scholars.

On the other hand, legal scholars, and comparativists in particular, tend to be interested in the comparative performance of alternative doctrinal solutions. Yet the typical comparative lawyer has no social science training. This will make it difficult for her to assess the probative value of comparing different outcomes in countries with different legal solutions for the same nor-mative problem. Of course, individual comparative lawyers could build this expertise. But this will at best take considerable time, and would divert their effort from what they are best at doing, i.e. the painstaking analysis of seemingly minute, but often highly relevant legal detail. Again using the metaphor, the translation of comparative legal work into data that lends itself to quantitative analysis, and the execution of this analysis, might be the comparative ad-vantage of social scientists interested in cross-country comparisons.

Hence seemingly there is ample room for mutually beneficial cross-interdisciplinary trade. Yet as others have documented, such trade is still conspicuously rare.[2] It would be a worthwhile exercise to search for "barriers to trade".[3] In the long run, institutional attempts at removing

---

[1] The theory has originally been developed in DAVID RICARDO, ON THE PRINCIPLES OF POLITICAL ECONOMY, AND TAXATION (J. Milligan; J. Gideon junior print. 1st American ed. 1819). For a modern treatment see AVINASH DIXIT & VICTOR NORMAN, THEORY OF INTERNATIONAL TRADE. A DUAL, GENERAL EQUILIBRIUM APPROACH (Cambridge University Press. 1980).

[2] Nuno Garoupa & Thomas Ulen, *Comparative Law and Economics. Aspirations and Hard Realities*, *** *** (2021);Holger Spamann, *Empirical Comparative Law*, 11 ANNUAL REVIEW OF LAW AND SOCIAL SCIENCE 131 (2015).

[3] For a repertoire of barriers to trade see European Union: Trade and Investment Barriers Report 2019, availa-ble at trade.ec.europa.eu/doclib/html/158789.htm (visited on Nov 25, 2020).

some of these barriers would be beneficial. Yet the aim of this article is more limited. I want to explore the scope of the enterprise.

I am bracketing theory. Learning about variance in legal institutions, or in legal discourse for that matter, may well inspire social scientists to develop new paradigms, and to generate testable propositions.[4] Yet for the purposes of this article, I assume that a social scientist has already generated such a proposition. She is intrigued by the observation that the legal rules, or legal practice when applying them, seem to differ across countries in ways that could help her put a theoretical claim to the test.

In this article I discuss why this seemingly obvious empirical approach is much harder than one might have thought.[5] Let me stress at the outset that I by no means want to contend that, therefore, the enterprise is futile, and comparative lawyers on the one hand, and social scientists on the other hand, should stay in their silos. Yet if the exchange between both camps is to gain momentum, the considerable methodological challenges should not be swept under the carpet. There is much to be gained from this collaboration. But the social science importers of comparative law must think hard about containing the risks, as must comparative lawyers if they intend to exploit the power of social science methods, and of quantitative analysis in particular. At the least, limitations must be acknowledged. And to the extent that the risks remain substantial, alternative research approaches are worth considering.

II.   Causal Claims

At their core, the empirical social sciences test causal claims. In the prototypical design, some theory predicts that some input causes some output. A classic illustration is positive economic theory,[6] or what political scientists and sociologists call the rational choice paradigm.[7] The theory is consequentialist.[8] It builds on the assumption that individuals choose between alternative actions anticipating their effects, and evaluating them according to the benefit they expect to accrue, net of the cost or risk involved. As law and economics has shown over decades, this is by no means the only, but in many contexts a thought-provoking way of analysing legal institutions.

To illustrate, one may be interested in the effect of liability rules on behaviour. A rational choice theorists would assume that the risk of being held liable is anticipated. If the expected loss

---

4   For one such attempt, see Joseph A. Conti, Relational Sociology and Comparative Law (2021).
5   The paper most similar in spirit, and equally cautious, is Spamann, ANNUAL REVIEW OF LAW AND SOCIAL SCIENCE, (2015).
6   For a comparative lawyer interested in better understanding the paradigm, law and economics theory would be a good starting point. At its core, the theory embraces the paradigm (maybe, in more recent years, allowing for behavioral qualifications). The introductory texts into law and economics are written with the intention to engage lawyers without modelling training. This makes them more accessible than treatises targeting economics students. And the paradigm is applied to legal applications. Good introductions include RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW (Aspen Publishers 6th ed. 2014);ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS (Prentice Hall 4th ed. 2012).
7   For a summary treatment, see RAFAEL WITTEK, et al., THE HANDBOOK OF RATIONAL CHOICE SOCIAL RESEARCH (Stanford University Press. 2013).
8   For the history and the conceptual foundations of consequentialism, see JULIA DRIVER, CONSEQUENTIALISM (Routledge. 2011).

resulting from having to compensate the victim exceeds the benefit from violating her, the constructed individual would refrain from the action.[9] Hence within the rational choice paradigm, a deterrence perspective on torts invites itself.[10] Actually the theory is even more precise. If the expected loss is smaller than the benefit, the theory would predict that the prospect of liability does not affect the choice of the activity level, or of precautionary measures. This expectation is itself a theoretical construct. In the simple most model, the theorist additionally assumes that the decision-maker doesn't worry about risk (is risk neutral). Then the amount of compensation can be multiplied with the probability of enforcement. This precision of the theory is desirable as it yields a clear prediction. There should be no difference in tortious behaviour if either (a) there is no liability or (b) the expected value of compensation is below the benefit of breaking the rule.

Now the tort law in different legal orders has embraced deterrence as one of the goals to a different degree.[11] In the US, punitive damages have a long history.[12] By contrast, the civil law jurisdictions have been much more hesitant to allow the amount of damages to exceed the harm suffered by the plaintiff.[13] A social scientist interested in testing the deterrence effect of tort law may therefore be tempted to compare some measure for tortuous behaviour between the US and, say, Germany.

This is when interaction with a comparative legal scholar would be highly beneficial. In the US, tort law is state law, and differs between states. In Germany, the general clause in § 823 I BGB is complemented by a host of more specific rules, scattered over multiple statutes.[14] Jurisprudence has allowed for exceptions from the general rule that compensation may not exceed the harm, for instance for copyright violations.[15] Not so rarely, legal orders use deeply different doctrinal paths to reach very similar decisions in similar cases.[16] Closely interacting with a comparative lawyer may therefore help the social scientist to not compare apples with oranges.

But unfortunately, the choice of the object of comparison is by far not the only hazardous part of the enterprise. Actually, using evidence from multiple jurisdictions to test a theoretical claim almost epitomises the panoply of concerns to be taken seriously when testing a causal

---

9   For a thorough treatment see URS SCHWEIZER, SPIELTHEORIE UND SCHULDRECHT (Mohr Siebeck. 2015).

10   The canonical model on deterrence is Gary Stanley Becker, *Crime and Punishment. An Economic Approach*, 76 JOURNAL OF POLITICAL ECONOMY 169 (1968).

11   More from André Tunc, *Introduction, in* INTERNATIONAL ENCYCLOPEDIA OF COMPARATIVE LAW XI: TORTS 154, (André Tunc ed. 1974). The empirics are debated, Gary Schwartz, *Reality in the Economics of Tort Law. Does Tort Law Really Deter?*, 42 UCLA LAW REVIEW 377 (1994);Gary T Schwartz, *Mixed Theories of Tort Law. Affirming Both Deterrence and Corrective Justice*, 75 TEXAS LAW REVIEW 1801 (1996).

12   See only Theodore Eisenberg, *Measuring the Deterrent Effect of Punitive Damages*, 87 GEORGETOWN LAW JOURNAL 347 (1998);Cass R. Sunstein, et al., *Assessing Punitive Damages*, 107 YALE LAW JOURNAL 2071 (1998).

13   HELMUT KOZIOL & VANESSA WILCOX, PUNITIVE DAMAGES. COMMON LAW AND CIVIL LAW PERSPECTIVES (Springer. 2009).

14   For a comparative treatment in English language see CEES VAN DAM, EUROPEAN TORT LAW (OUP Oxford. 2013).

15   Bundesgerichtshof March 10, 1972, BGHZ 59, 286.

16   A classic is the doctrine of consideration. As such, it is not part of civil law. Yet civil law does not hold a promise against the promisor if it is classified as a gift, and common law complements the doctrine with promissory estoppel. As a consequence, there are only very few cases (like promising to pay an unknown person for bringing back a lost dog) where outcomes would differ. For detail see DIETRICH ROTHOEFT, SYSTEM DER IRRTUMSLEHRE ALS METHODENFRAGE DER RECHTSVERGLEICHUNG. DARGESTELLT AM DEUTSCHEN UND ENGLISCHEN VERTRAGSRECHT (Mohr <Siebeck>. 1968);FERDINAND FROMHOLZER, CONSIDERATION (Mohr (Siebeck). 1997)..

claim.[17] In the following I will make the reader understood why the enterprise is fraught with risk. Some of these risks are routinely discussed in the social sciences, and should therefore immediately induce any social scientist to become cautious when using a difference in legal rules across jurisdictions as a source of variance. Other concerns are more specific to this very empirical strategy. Unfortunately, these more specific concerns carry even more weight.

Every empiricist worries about **reverse causality**. Her theory predicts that A causes B. But for the most part all she is able to observe is the cooccurrence of A and B. If there is a lot of As, there are also a lot of Bs. But correlation is not causation. The fact that there are so many Bs may be exogenous, and the presence of the Bs may actually have engendered many As. Let me use the tort example for illustration. Under Sec. 4 Clayton Act,[18] a private party who has been injured in her business by a violation of anti-trust rules may sue the violator for treble damages, i.e. three times the harm she has suffered. The German Gesetz gegen Wettbewerbsbeschränkungen has no comparable rule. The probability of being sued for an anti-trust violation is certainly substantially below 100%. One might therefore expect that German law underdeters.

Let's assume, for the sake of the argument, that some clever choice of datasets is in line with the theoretical prediction. Does that prove the hypothesis? Not necessarily. It could be that there have been less anti-trust violations in the first place, so that the need for deterrence was less pronounced. This is another instance where collaboration with a comparative lawyer would be beneficial. This lawyer would probably suggest that the alternative explanation is unlikely to be true, given the Clayton Act is much older than the pertinent provisions under German law. But another channel of reverse causality may be harder to dismiss. Is has been argued that the US is a more litigant society than Germany. Shareholder value is the dominant concept in corporate law.[19] By contrast in Germany, the corporation has always been understood as an enterprise that has to serve shareholders and stakeholders alike, including the workforce and the physical environment.[20] Hence business in the US might have been more in need of deterrence than comparable business in Germany.

The worry about **omitted variables** looms equally large. A graphic illustration stems from a different domain. In the old days, children were made to believe that babies were brought by the stork. German data seems to provide evidence. Storks are much more frequent in the eastern part of the country.[21] This is also where the birth rate is highest.[22] Not a mystery, but an omitted variable. Before 1989, the eastern parts of the country were under Communist rule. As communism was substantially less successful economically, storks had an easier time finding

---

[17]   Excellent discussion in JOSHUA D ANGRIST & JÖRN-STEFFEN PISCHKE, MOSTLY HARMLESS ECONOMETRICS. AN EMPIRICIST'S COMPANION (Princeton University Press. 2008).

[18]   15 U.S.C.S. § 15.

[19]   Luca Enriques, et al., *The Basic Governance Structure. Minority Shareholders and Non-shareholder Constituencies*, *in* THE ANATOMY OF CORPORATE LAW. A COMPARATIVE AND FUNCTIONAL APPROACH (Reinier Kraakman ed. 2017).

[20]   German Corporate Governance Code, https://www.dcgk.de//files/dcgk/usercontent/en/download/code/191216_German_Corporate_Governance_Code.pdf, (visited Nov 26, 2020). https://bergenhusen.nabu.de/imperia/md/nabu/images/nabu/einrichtungen/bergenhusen/germany__zensus_2004.pdf. 2004 all over Europe, white storks have been counted, https://bergenhusen.nabu.de/weissstorch/bestand-international.html, (links visited Nov 27, 2020).

[22]   https://www.bib.bund.de/DE/Fakten/Regional/Geburten-Sterblichkeit.html (visited Nov 27, 2020).

a habitat. On the other hand, the ideology of the ruling party put a high value on women being part of the labour force. This led to considerably better childcare, encouraging more women to have children.

Omitted variable bias has also to be taken seriously when using a difference in some economic or societal outcome between two countries as evidence that a difference in law between the countries has been causal. No two countries are the same. Often, some of these differences can be observed. One for instance has information about the distribution of income, or about political stability. If one has the information, one can condition the effect of interest, i.e. of a difference in legal rules, on these other, potentially relevant factors.[23] But it is not at all rare that one suspects further influences, but has no data on them. Then the difference in outcomes cannot be taken as support for the theoretical claim.

Let me again illustrate the concern with the question whether the obligation to compensate victims deters firms from breaking anti-trust law. Let us assume that the difference in outcomes between the US and Germany is too small to be meaningful.[24] Can the researcher infer that deterrence is not important for anti-trust law? The prospect of having to compensate the victim would be non-deterrent if (a) there is at least a small probability that victims will not sue, or that they might lose in court, and (b) the harm suffered by the suing victims is no larger than the benefit from breaking the law. Then in expectation, the firm still makes at least a small profit. If the firm maximizes profit, and exclusively cares about the risk of litigation, it breaks the law. Yet private litigation is not the only intervention to enforce anti-trust rules. There is also public enforcement by the anti-trust authorities, and the violation of the rules can constitute a crime. The overall deterrent effect results from a cumulation of these enforcement activities.[25] Again the social scientist would be well advised to collaborate with a legal scholar who could inform her about these additional enforcement channels. And it would be necessary to quantify the deterrent effect achieved on these other channels to find out whether there is a residual effect of the prospect of private litigation.

A related concern is **selection**. The researcher wants to make a statement about the effect in the population. She, however, usually has only access to a sample. If this sample is randomly selected from the population, and sufficiently large, the researcher may be confident that the effect in the sample pictures the effect in the population at large. Yet this confidence breaks down if the sample is not a representation of the population. The researcher is then led to believe in an effect, or the absence of an effect for that matter. But she actually can only make a statement about the subgroup of the population that she had a chance to study. Quite often the latter statement is not of interest. Actually, selection can be interpreted as omitted variable

---

[23]   In statistical jargon: one can „control for" these alternative explanations.

[24]   Does not meet the conventional level for statistical significance. In the social sciences, this level is conventionally set at 5%. Technically, this is the probability of obtaining test results at least as extreme when redrawing an infinite number of samples from the same population, provided the null hypothesis is true. Hence this p-value defines the risk of a false positive result (one rejects the null hypothesis, although there is no support for the alternative hypothesis in the population).

[25]   For the US see Robert H Lande & Joshua P Davis, *Comparative Deterrence from Private Enforcement and Criminal Enforcement of the US Antitrust Laws*, Brigham Young University Law Review 315 (2011).

bias. The researcher has neglected to control for the fact that her sample has specific characteristics.

Selection is pervasive if data generated by the judiciary is used as evidence.[26] People do not sue at random. Among the already selective set of cases that are sued in the first place, those that make it to superior courts are again not randomly selected. Frequently, only the latter are reported and can be used as evidence. Litigation is strategic, on the part of the parties,[27] and sometimes even on the part of the courts.[28]

Selection is particularly concerning if the researcher wants to compare across jurisdictions. Jurisdictions have very different reporting standards.[29] What looks like a difference in substantive outcomes, and therefore like support for the hypothesis to be tested, may actually just result from different reporting cultures. In this respect, support by a comparative lawyer is particularly urgent. This lawyer has to have a fairly deep understanding of the judicial systems to be compared to assess the biasing effect of reporting cultures. Merely comparing the available decisions along reported dimensions will not suffice. Even if the available body of cases looks sufficiently similar in the jurisdictions to be compared, the rulings may have made it into the reports for very different reasons. One for instance has to know that the judges in German higher courts used to earn additional money for handing in decisions to commercial reporters. They thus had a business interest in the composition of the body of published cases.

In its bluntest form, **measurement error** has been featuring prominently in the scholarly debate over differences in the law as a cause of differences in economic performance. The "legal origins" literature has claimed that common law is more responsive to the needs of the economy, has therefore better protected shareholders, which has helped the economies in common law countries to grow more steadily.[30] The underlying coding of legal orders had been undertaken without much access to legal expertise. Once trained comparative lawyers stepped in, and redid the coding properly, the claimed effect largely vanished.[31]

Now it stands to reason that one cannot come to unbiased conclusions on the basis of biased data. It is less obvious why data analysts should even be concerned if there are no signs of bias. This is the situation that the statistical literature refers to when talking about measurement error.[32] The explanatory variable of interest can only be measured imprecisely. More

---

[26]   George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 THE JOURNAL OF LEGAL STUDIES 1 (1984).

[27]   Michael Ramsden & Kris Gledhill, *Defining Strategic Litigation*, 4 CIVIL JUSTICE QUARTERLY 407 (2019).

[28]   Daniel Klerman & Greg Reilly, *Forum Selling*, 89 SOUTHERN CALIFORNIA LAW REVIEW 241 (2015);Stefan Bechtold, et al., *Forum Selling Abroad*, 92 SOUTERN CALIFORNIA LAW REVIEW 487 (2018).

[29]   Cf. Peter Harris, *Ecology and Culture in the Communication of Precedent Among State Supreme Courts, 1870-1970*, LAW AND SOCIETY REVIEW 449 (1985).

[30]   Rafael La Porta, et al., *The Economic Consequences of Legal Origins*, 46 JOURNAL OF ECONOMIC LITERATURE 285 (2008).; for an extensive review of this literature, see Garoupa & Ulen, ***, (2021).

[31]   John Armour, et al., *How Do Legal Rules Evolve? Evidence from a Crosscountry Comparison of Shareholder, Creditor, and Worker Protection*, 57 AMERICAN JOURNAL OF COMPARATIVE LAW 579 (2009);John Armour, et al., *Shareholder Protection and Stock Market Development. An Empirical Test of the Legal Origins Hypothesis*, 6 JOURNAL OF EMPIRICAL LEGAL STUDIES 343 (2009). See also the critique by Ralf Michaels, *Comparative Law by Numbers? Legal Origins Thesis, Doing Business Reports, and the Silence of Traditional Comparative Law*, 57 AMERICAN JOURNAL OF COMPARATIVE LAW 765 (2009).

[32]   See Jonathan Gillard, *An Overview of Linear Structural Models in Errors in Variables Regression*, 8 REVSTAT–STATISTICAL JOURNAL 57 (2010).

technically, the initial concern is noise, not bias. Now statistical evidence routinely comes with a measure of confidence. In principle, if the explanatory variable can only be measured with error, the job of the statistician merely becomes harder. The true effect must be bigger to meet the conventional criterion for confidence (the significance level). Seemingly, not much harm can be done. If measurement is too noisy, the evidence does not support the theoretical claim. Empiricists typically don't worry too much about being too conservative.

But there is a catch. Very often statistical analysis is not confined to assessing the effect of one explanatory variable on one outcome variable. Rather one simultaneously assesses the explanatory power of multiple variables. This approach is particularly appealing in testing for a purported effect of the law. For with this procedure one may run a horse race between alternative (legal or extralegal) explanations. Now in such richer statistical models the fact that one candidate explanation is only measured with error can indeed cause bias. As the first, noisy explanation does not succeed in explaining the variation in the data, the alternative explanations carry more weight, and may well become significant. The analyst is led to believe that the data supports the alternative explanation although it would have remained insignificant, had the first explanation been measured cleanly.

In a way, a comparative lawyer always compares apples with oranges. It will rarely take a lot of ingenuity to spot differences between the objects of comparison, i.e. the legal orders, that transcend the difference of interest. Ultimately to prove causality, one would need the counterfactual. It is unknown by definition. In the language used in the statistical literature, one and the same case cannot simultaneously be "treated" and remain "untreated".[33] The standard response of the empirical literature is not available when testing the effect of alternative legal solutions with the help of data from different jurisdictions. Legal rules are not **randomly assigned** to a random selection from the pool of jurisdictions.

An imperfect, but appealing way out is adding a second dimension to the data. Rather than only considering outcomes from two (or more) jurisdictions, one observes the outcome of interest in all jurisdictions over a protracted amount of time. In statistical parlance, one then goes beyond testing a "cross-section", and exploits the information about development over time in a "panel".[34] This empirical strategy is particularly appealing if, during the period of observation, the rule of interest has changed in one jurisdiction, but not in the other. If the outcome variable changes in the "treated" jurisdiction while it remains constant in the "untreated" jurisdiction, one may be reasonably confident that the change has indeed been caused by the intervention, i.e. by the entry into force of the rule in question.

Yet even such a "**difference in differences**" design has its limitations.[35] The design does not work in the first place if, before the rule has changed in the "treated" jurisdiction, the outcome variable has exhibited a different development across the investigated jurisdictions. It is this "common trend" before intervention that makes the approach credible. Yet even if the common trend can be established, one should be cautious with interpretation. What one sees is the

---

33   The logic is powerfully explained in ANGRIST & PISCHKE. 2008.
34   For background see, e.g., MARNO VERBEEK, A GUIDE TO MODERN ECONOMETRICS (John Wiley & Sons. 2008).
35   More from Richard Blundell & Monica Costa Dias, *Alternative Approaches to Evaluation in Empirical Microeconomics*, 44 JOURNAL OF HUMAN RESOURCES 565 (2009).

effect of a change in rules. Yet the question of normative interest usually refers to the presence or the absence of a rule.[36] Anecdotal evidence suggests that this difference matters in the legal domain. The concern is, for instance, known from criminal law. Legislators have often reacted with making punishment more severe to the perception that crime is on the rise. Even if, in the short run, the incidence of the crime in question goes down, it may stabilise at the original high level after the community of criminals has gotten used to the more severe rule.[37] The effect of changing the rule has not been sustainable.

Actually the reality of legal reform frequently runs counter being useful for estimating a causal effect in even deeper ways. Legal rules do not happen to change. They are changed for a purpose. Quite often this purpose is not merely an ideological difference between the previous and the present government. The change in rules has been triggered by the perception of a social ill. The amended rule is expected to address the ill. From the perspective of a social scientist this constitutes reverse causality. The concern is most obvious if the jurisdiction that has changed the rule had been more deeply affected by the normative problem, or at least had been more sensitive to it. Then the treated and the untreated jurisdictions are no longer comparable. Assessing this source of bias requires fine-grained contextual knowledge, which a comparative lawyer will be much more competent to muster than the social scientist.

Moreover legal rules are not designed, or changed for that matter, by a benevolent dictator. In a democracy, legislation is regulative policy-making.[38] Legal theorists may draw a strict line between rule generation and rule application. It marks the difference between a debate *de lege ferenda* and *de lege lata*. If this divide can be taken for granted, the social scientist interested in gauging the effect of the rule need not bother. But in many contexts one has reason to be sceptical. Some rules remain law in the books forever.[39] Others are designed with the intention to make implementation difficult,[40] for instance since the legislator wants to engage in merely symbolic policy-making.[41] Yet other rules require that the law's subjects avail themselves of the new opportunity.[42] Any of these qualifications makes it difficult to precisely define from which point in time on, and to which degree, the outcome variable of interest would have to change if the theoretical claim of a causal effect is well-founded. This is another instance where a comparative lawyer, with full command not only of legal doctrine, but also of implementation practice, in all jurisdictions to be compared, is needed.

All of the foregoing should be obvious for a well-trained empirical social scientist. After all the social sciences have undergone their identification revolution.[43] Every grad school teaches

---

[36]   For an experimental demonstration of the difference see Christoph Engel, *A Random Shock is not Random Assignment*, 145 ECONOMICS LETTERS 45 (2016).

[37]   Anthony N Doob & Cheryl Marie Webster, *Sentence Severity and Crime. Accepting the Null Hypothesis*, 30 CRIME AND JUSTICE 143 (2003).

[38]   ROLAND CZADA, et al., REGULATIVE POLITIK. ZÄHMUNGEN VON MARKT UND TECHNIK (Springer-Verlag. 2013).

[39]   Roscoe Pound, *Law in Books and Law in Action*, 44 AMERICAN LAW REVIEW 12 (1910).

[40]   Renate Mayntz, *Implementation von regulativer Politik*, *in* IMPLEMENTATION POLITISCHER PROGRAMME II (Renate Mayntz ed. 1983).

[41]   MURRAY J. EDELMAN, THE SYMBOLIC USES OF POLITICS (University of Illinois Press. 1964).

[42]   Christoph Engel, et al., *Diffusion of Legal Innovations: The Case of Israeli Class Actions*, 15 JOURNAL OF EMPIRICAL LEGAL STUDIES 708 (2018). demonstrate that this cannot be taken for granted, in the case of class action in Israel.

[43]   Edward E. Leamer, *Let's Take the Con out of Econometrics*, 23 AMERICAN ECONOMIC REVIEW 31 (1983).

that correlation is not causation, and prepares the next generation of social scientists for ideally not only discerning and acknowledging the limitations of the analysis. They also help them creating designs that contain the concern.[44] One powerful way of mitigating the problem is an indirect one.

One can describe an identification problem as a problem of endogeneity.[45] Ideally one would want to cleanly separate cause and effect. Now for social reality it would not be meaningful to postulate natural laws. Social life is too complex to be treated like gravity. Social scientists take it for granted that they can only observe the causal relationship of interest with a certain degree of noise. The gold standard is not a perfect mapping between both of them. A social scientist accepts a theoretical claim, at least for the time being, if it is sufficiently unlikely that the difference in outcomes between treated and untreated observations results from noise. She thus accepts residual error. Endogeneity is present if the unexplained part of the data is actually not just noisy, but informative about the explanatory variable.

This is where the solution starts. One may be in the position to replace the contaminated explanatory variable with another, sufficiently clean variable.[46] This strategy helps if the replacement, in statistical jargon called an **instrument**, is sufficiently correlated with the contaminated variable that one is interested in. Of course, unless the correlation is near-perfect, one loses explanatory power. One essentially only has the correlation between the two variables, rather than the variable of interest. But this is the price one has to pay for removing the potential bias resulting from endogeneity.

Technical as this may sound, this is one of the most productive ways for the interdisciplinary interaction between comparative lawyers and social scientists. Finding good instruments is conspicuously hard. Either the instrument is actually also correlated with the error term, and then not valid. Or the instrument is truly exogenous, but the correlation between the instrument and the potentially contaminated explanatory variable of interest is too low, which is why the instrument is weak, and therefore useless.[47] In the social sciences researchers are rightly rewarded with publications in the best journals if they can come up with a convincing instrument that is neither endogenous nor weak.

Once a comparative lawyer has understood what is needed for proper instrumentation, she can bring her knowledge to bear. Precisely because the discipline excels in the painstaking analysis of minute detail, it is so precious for the social scientist. Actually at the core of the legal origins literature discussed above there is one such clever idea. Whether a country can be classified as a common law or a civil law jurisdiction has been decided a long time ago. In the Middle Ages in Europe, by colonial powers in many other countries of the world. To the degree that true differences between common law and civil law persist to these days, one has

---

[44]   Joshua D. Angrist & Jörn-Steffen Pischke, *The Credibility Revolution in Empirical Economics. How Better Research Design is Taking the Con out of Econometrics*, 24 JOURNAL OF ECONOMIC PERSPECTIVES 3 (2010).

[45]   For background see ANGRIST & PISCHKE. 2008.

[46]   For a very accessible introduction see Joshua D Angrist, *Instrumental Variables Methods in Experimental Criminological Research. What, Why and How*, 2 JOURNAL OF EXPERIMENTAL CRIMINOLOGY 23 (2006).

[47]   John Bound, et al., *Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak*, 90 JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 443 (1995).

indeed found an instrument. One may use the correlation between legal origin and one of the modern-day legal institutions to instrument the latter. Hardly any topic has been dissected that thoroughly by comparative lawyers as the differences, seeming and real, between common law and civil law. This equips them to detect powerful instruments that a social scientist would have a hard time finding on her own.

In the tradition of comparative law, there is great respect for detail. Even if, at the surface, the solutions found by two legal orders seem to differ profoundly, putting these doctrinal constructs into context may well reveal that the differences are much smaller, if any. And reversely the fact that one legal order seems to use the same doctrinal technique as another is not sufficient to conclude that the two legal orders agree on the solution. The latter possibility looms particularly large in legal rules inherited from a colonial past,[48] or introduced by way of transplant.[49]

From the perspective of quantitative analysis, the resulting concern goes beyond misclassification. The very nature of the exercise is called into question. Statistical analysis requires a **population**. In the respect of interest, one must be able to argue that the population is homogeneous. Then observing a sufficiently large sample, once treated and once untreated, makes it possible to infer how other entities in the population react to the intervention in question. If actually every entity is different from each other, there is no room for inference "out of sample". Now every legal order is a product of history. History does not repeat, neither within nor across countries. Seemingly the standard critique of (comparative) lawyers can therefore even be substantiated in the language of statistics. Seemingly, comparative law is just not an appropriate object of quantitative analysis.

Yet this radical statement throws out the baby with the bathwater. Statistical analysis has been productively employed for the analysis of individual behaviour, and of firms in the market. Now each individual has its idiosyncratic history, as has each firm. Belabouring a term that will be more familiar to comparative lawyers, the critical issue is abstraction. The statistical concept of a population is a construct. The empiricist (in practice implicitly) defines this construct when specifying the empirical model. As with theory, one has to acknowledge that, as it has been famously argued, all models are wrong, but some are useful.[50] Comparative lawyers should therefore not invest in demonstrating that the respective population is constructed; it always is. They should instead discuss whether the features of individual legal orders from which the respective empirical model abstracts leave out elements that may not be neglected for the resulting evidence to be meaningful.

Another feature of comparative law is more concerning. For the interaction between social scientists and comparative lawyers to be productive, some outcome variable of interest must be compared at the level of legal orders. At first glance, a strong case for quantitative analysis can be made. There are more than 200 sovereign states. All of them have a legal order; federal

---

[48]   See for instance Maryam Kanna, *Furthering Decolonization. Judicial Review of Colonial Criminal Law*, 70, DUKE LAW JOURNAL 411 (2020).
[49]   See for instance Holger Spamann, *Contemporary Legal Transplants. Legal Families and the Diffusion of (Corporate) Law*, BRIGHAM YOUNG UNIVERSITY LAW REVIEW 1813 (2009).
[50]   George EP Box, *Science and Statistics*, 71, *Journal of the American Statistical Association,* 791 (1976).

states do even have multiple. Usually, statisticians feel quite comfortable with a few hundred observations. Most of the tests they are using to weed out spurious relationships require asymptotic theory. The tests rely on distributional assumptions. But by the central limit theorem, with sufficiently many observations the concern vanishes.[51] The distribution mechanically approximates normality. If normality can be assumed, the standard statistical procedures for estimating central tendencies and standard errors are reliable.

Yet this line of argument requires sufficiently many independent observations. If 100 randomly selected individuals receive some intervention, while another hundred individuals randomly selected from the same population are left without intervention, and if outcomes of the former differ systematically and sufficiently strongly from outcomes of the latter, one may confidently infer an effect at the population level. Yet if the intervention in question is the reaction to the severity of punishing robbery, and the treated group is a cloister of monks, while the control group is a gang, one would not expect to draw a proper image of the intervention at the level of the population at large. Clerical life impacts on criminal behaviour, as does gang membership.[52]

While every empirical researcher would be alerted in such a striking case, dependence problems can be more subtle, and still be concerning. The most important source of dependence in the case of comparative law is what the empirical literature would call contamination.[53] Before deciding upon the act that is put to the statistical test, the entity in question has come under the influence of some other entity in the sample. For the very reason that legal orders are products of history, this is likely to be the case between jurisdictions. Legal rules are normally not designed from scratch. The standard *modus operandi* of the legislator is not design, but reform. As comparative lawyers have very elegantly worked out, the common origin, for instance in common or Roman law, has left discernible traces on many legal orders.[54]

Moreover, legislators observe each other, and developments in the jurisprudence of other countries.[55] Sometimes they adopt *en bloc* what other jurisdictions have had previously. The spread of anti-trust legislation throughout the world, starting with the US, the EU and Germany, is a graphic illustration.[56] In other instances, the legislator of one country does not simply copy

---

[51]   Imre Bárány & Van Vu, *Central Limit Theorems for Gaussian Polytopes*, 35 Annals of Probability 1593 (2007).

[52]   This ties into the critique of universalism proffered by Tom R. Tyler, The role of comparative law in the social sciences. Understanding the psychology of social order. (2021).

[53]   The topic is most frequently discussed with respect to repeated observations from groups of interacting individuals. From the second period on, choices are no longer independent. The classic reaction of statisticians is estimating a mixed effects model. In such a model, the dependence is explicitly modelled, in the most straightforward case by a random effect at the level of individuals, crossed with a further random effect at the level of groups. This specification is appropriate if one may argue that all group level effects are captured by the latter effect (the random effects are orthogonal), and that the group and individual effects are uncorrelated with the explanatory variables (there is no endogeneity). The approach, however, only works if one has sufficiently many, truly independent, units at the highest level of dependence. Otherwise the requirements from asymptotic theory for inference from the sample to the population are not met, A Colin Cameron, et al., *Bootstrap-based Improvements for Inference with Clustered Errors*, 90 Review of Economics and Statistics 414 (2008).

[54]   See, for instance, Konrad Zweigert & Hein Kötz, Introduction to Comparative Law (Clarendon Press 3rd rev. ed. 1998).

[55]   See, e.g., Maximo Langer, *From Legal Transplants to Legal Translations. The Globalization of Plea Bargaining and the Americanization Thesis in Criminal Procedure*, 45 Harvard International Law Journal 1 (2004).

[56]   Keith N. Hylton & Fei Deng, *Antitrust Around the World: An Empirical Analysis of the Scope of Competition Laws and Their Effects'(2007)*, 74 Antitrust Law Journal 271 (2007).

the solution of another. But when discussing their own intervention, it benefits from the experiences other jurisdictions have made.[57] If the country in question is a democracy, the interplay between government and opposition is a further channel of influence. Either of them will point to foreign solutions that seem more conducive to further their own political convictions.[58] Yet another channel is business. Even if, originally, a country desires to protect its legislative autonomy, pressure from business actors may force them to align their solution to a prominent legislative trend.[59] As the debate over a California[60] or Brussels effect[61] has demonstrated, such pressure may even ultimately help establish the more stringent, not the more lenient standard.

From the perspective of quantitative analysis, such cross-fertilisation creates a challenge. Unless one may confidently argue that the foreign impulses have been negligible, the observations from different countries, or jurisdictions for that matter, are no longer independent from each other. As has been explained, this invalidates the established procedures for assessing confidence in quantitative results.

Note that the concern does not go away if the outcome variable is not legal in nature. One may, for instance, be interested in the effect of different inheritance rules on the birthrate, or in the effect of different techniques for consumer protection on the development of online trade. With such research questions, cross-fertilisation between legal orders does not have a direct effect on the extra-legal dependent variable. Yet the project is motivated by the hypothesis that outcomes are, at least partly, determined by differences in legal rules. If, in turn, one suspects these rules to be influenced by each other, on this indirect channel there is dependence.

III. Solutions

No evidence whatsoever is almost certainly worse than evidence with limited probative value. The foregoing should therefore not be misread. Comparative law can benefit from quantitative methods, as can quantitative social science from the interaction with comparative lawyers. All one should derive from the previous section is a call for modesty.[62] Observing some outcome in two jurisdictions that differ with respect to legal solutions is not a panacea.

---

[57] Consumer protection is a pertinent illustration, Stephen Weatherill, *Consumer protection*, *in* ELGAR ENCYCLOPEDIA OF COMPARATIVE LAW, SECOND EDITION (2012).

[58] A good illustration is the European debate over the desirability of class action, JÜRGEN G BACKHAUS, et al., THE LAW AND ECONOMICS OF CLASS ACTIONS IN EUROPE. LESSONS FROM AMERICA (Edward Elgar Publishing. 2012).

[59] See, for illustration, the debate over accounting standards, Ann Jorissen, et al., *A Geographic Analysis of Constituents' Formal Participation in the Process of International Accounting Standard Setting. Do we have a Level Playing Field?*, 32 JOURNAL OF ACCOUNTING AND PUBLIC POLICY 237 (2013).

[60] DAVID VOGEL, TRADING UP. CONSUMER AND ENVIRONMENTAL REGULATION IN A GLOBAL ECONOMY (Harvard University Press. 1995).

[61] ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (Oxford University Press, USA. 2020).

[62] In this respect, comparative law is just one branch of legal scholarship. The law engages with the lives of real people, and can therefore usually not afford to artificially simplify the research question, and the data used for responding, in the interest of constructing a well-defined problem. At their core, legal problems are ill-defined. This forces legal scholars to face the dilemma between doing justice to the normative problem at hand, and scientific standards, in ways conventionally defined away by the social sciences. More on this

One option that always exists is making the limitations explicit. Depending on the research question, the possibility of reverse causality may seem far-fetched, and measurement error may be unlikely as well. Probably dependence problems would still persist, and the constructed population may still appear somewhat artificial. But the more the result resonates with plausible theory, the more one may feel justified further pursuing this line of investigation. Social scientists would call such projects descriptive, or exploratory. Strictly speaking, the results would only generate, not test hypotheses. But this may have value, in particular in directing further research.

Moreover, not each and every research question requires comparing outcomes at the macro, i.e. at the jurisdiction level. If, for instance, one is interested in some determinant of judicial decision-making in the lower courts, say the degree of professional experience required for assuming the office, one may try to construct a carefully selected data-set from two jurisdictions that differ in this respect, and may try to hold as many alternative explanations constant as possible.[63] This may in principle be done with the help of stratifying the sample in multiple dimensions. Usually this approach fails, however, as cell sizes get too small: there are two few individual judges with the exact same combination of parameters in multiple dimensions. This is when social scientists reweight the data,[64] to construct a synthetic control group.[65] Note, however, that this method can only get under control what is observed. The researcher must be able to argue that further, unobserved differences are negligible. And by design the method cannot control for macro-level differences: all "treated" observations come from a jurisdiction with the same legal order, the same legal culture, the same political system, the same economic climate, and so forth.

There are two more strategies worth mentioning. The core of the identification problem described above is the nature of the data. The researcher is interested in the effect of some legal solution on some (legal or extra-legal) outcome variable. She directly tests her theoretical claim with the object of inquiry. Now as I have explained above, all quantitative analysis is constructive. The specification of the statistical test relies on a model. No model is a complete mapping of some segment of reality. This is not a bug, but a feature. Only because the analyst zeros in on the channel of interest is she able to be falsified. Falsification is, at its core, how quantitative analysis differs from storytelling.[66]

Once one has understood the selective nature of the enterprise, it will become more palpable to already replace the data. Rather than studying the relationship of interest with the data of interest, the researcher may focus on the former. Her theory tells her that some legal rule should have a defined effect on some outcome variable. This outcome will frequently be a

from Christoph Engel, *Empirical Methods for the Law*, 174 Journal of Institutional and Theoretical Economics 5 (2018).

[63] Lee Epstein, et al., The Role of Comparative Law in the Analysis of Judicial Behavior (2021).

[64] Alberto Abadie & Guido W Imbens, *Matching on the Estimated Propensity Score*, 84 Econometrica 781 (2016).

[65] Alberto Abadie, et al., *Comparative Politics and the Synthetic Control Method*, 59 American Journal of Political Science 495 (2015).

[66] Karl Popper, The Logic of Scientific Discovery (Routledge. 2005).

decision by some discernible entity, say an individual. Then one may create new data that directly tests the purported causal effect.

The social sciences have developed a whole panoply of methods for the purpose.[67] They differ by the degree of context that they require, or accept for that matter. On the opposite end of the methodological spectrum is the laboratory experiment. In the interest of isolating the channel of influence, all context is deliberately stripped away. The typical experiment with a legal research question does not even talk about the ultimate goal to cast light on the effect of the legal rule in question. Rather the experiment is narrowed down to the one feature of the legal institution that theory predicts to do the trick. In that spirit, one may, for instance, want to understand whether the fact that judges are elected, rather than appointed, creates bias.[68]

Ultimately the researcher faces a trade-off between internal validity (the relationship between cause and effect is beyond doubt) and external validity (the object of investigation is identical with, or at least very similar to, the object to which the theory refers).[69] There is an intermediate zone between maximum internal validity (laboratory experiment) and maximum external validity (observational data). In this set of intermediate methods, the researcher is willing to pay a price in terms of residual doubt about identification, in the interest of getting closer to the object of the inquiry. Or, the other way around, the researcher is willing to pay a price in terms of studying a phenomenon that is only analogous to the phenomenon of interest, in the interest of being more confident about causality. Field experiments put more stress on internal validity. There is still random assignment to treatment. But treatment is under real-life conditions, which can never be standardised to the degree this can be done in the laboratory. Vignette studies put more stress on external validity. Participants are randomly exposed to context rich scenarios, and are asked to give a hypothetical answer. The data is less credible though, as responses are only hypothetical. But scenarios can otherwise be very realistic, and the method is also available if field experiments would be impractical or illegal. This will hold true for many research questions of interest when comparing jurisdictions. Take, for instance, family planning. No internal review board would clear a field experiment that randomly exposes families to contexts that make it less or more likely for them to want children. Asking hypothetically how they would react if a certain rule were in place will be much more acceptable.

Finally, quantitative methods are not confined to the test of causal claims. The researcher may adopt an alternative epistemic goal. Actually, this research strategy has become very prominent lately, through the availability of large legal datasets, and high computing power. This is where comparative law may productively interact with machine learning.[70]

---

[67]   Simeon Yates, Doing Social Science Research (Sage. 2003).
[68]   Christoph Engel & Lilia Zhurakhovska, *You Are In Charge. Experimentally Testing the Motivating Power of Holding a Judicial Office*, 46 Journal of Legal Studies 1 (2017).
[69]   Brian E Roe & David R Just, *Internal and External validity in Economics Research. Tradeoffs Between Experiments, Field experiments, Natural Experiments, and Field Data*, 91 American Journal of Agricultural Economics 1266 (2009).
[70]   A very accessible introduction to the methods employed is Gareth James, et al., An Introduction to Statistical Learning § 112 (Springer. 2013).

Computer scientists have developed a whole array of methods for organising rich data. These methods are designed to detect patterns. A classic application in the area of comparative law is the claim that there are legal families. This claim can be directly translated into computer science. If the claim holds true, then it should be possible to find discernible patterns in the data from multiple jurisdictions. These patterns could be found at the level of rules, but also at the level of decisions taken on the basis of these rules. Actually one can push the approach even a step further. Comparative law does not only claim that legal rules, and legal practice for that matter, exhibit systematic relationships among each other. Comparative law even pinpoints specific legal families. One may use the power of clustering algorithms to try and reconstruct these families. If the set of clusters found with the help of the computer differs from the comparative law tradition, this is the thought-provoking starting point for joint work between social scientists and comparative lawyers.[71]

Conclusion

Causal claims are tempting. Legal scholars would want to know whether a proposed reform would likely deliver on its promises. Causal claims are equally relevant in a backward looking perspective. A classic application is the principle of proportionality. If constitutional jurisprudence adopts the principle, a legal rule will only be upheld if the case can be made that it serves some purpose of social betterment. The rule must be the cause for a state of social affairs that is preferable over the state of affairs to be expected should the rule be invalidated. Political scientists would call this an investigation into the effectiveness of regulative politics. Economists may wonder whether a society governed by a certain rule effectively contains a clash between individual and social rationality. Sociologists may argue that a society embedded into a certain set of formal legal rules is more resilient to shocks than another. Psychologists may investigate whether the presence of some legal rule shapes the process by which individuals acquire information about some societal phenomenon of interest, say polarization around some leader. Jurisdictions are sovereign. This is why legal rules differ between jurisdictions. In terms of research questions, there is thus ample room for the productive collaboration of and cross-fertilization between comparative lawyers and social scientists.

Yet testing causal claims is hard. The ideal setting is impossible by definition. One can never observe one and the same social setting, once with the respective legal rule in force, and once in its absence. The perfect counterfactual is imaginary. The social sciences have developed a series of proxies, and conventions about their acceptability. Unfortunately, comparative law is a particularly challenging type of evidence for the use of these proxies. These challenges have been worked out extensively by the empirical social sciences. For some challenges, in appropriate settings, the precise command that comparative lawyers tend to have of doctrinal detail may help construct clever workarounds. But in particular the problem of dependence is almost insurmountable. Legal rules do not develop independently of legal rules developed in other jurisdictions.

---

[71] Yun-chien Chang, et al., *Drawing the Legal Family Tree: An Empirical Comparative Study of 108 Property Doctrines in 128 Jurisdictions*, JOURNAL OF LEGAL ANALYSIS (2020).

If the identification of a causal effect is not attainable, weaker empirical designs may still be feasible. One popular option is the construction of a synthetic control group (which, however, can only take known and understood intervening variables out of the equation). A safer option is the generation of new data, using some experimental method. Then identification can be achieved with the help of random assignment to treatment. Yet jurisdictions are not commonly willing to experiment with a rule borrowed from a different jurisdiction. Legislatures will not do that lightly. Vignette studies or lab experiments are easier to implement. But then what one studies is only analogous to what one wants to understand.

This paper should not be abused as a plea against the use of social science methods in comparative law, or against the use of comparative law as evidence in the social sciences. This collaboration has high potential. There is merely no one size fits all design. All empirical methods have limitations. The researcher must choose carefully, and must explicitly motivate why the limitations inherent in the chosen method are less concerning than the limitations when using alternative approaches. Not even trying, and just mustering common sense, is not likely to be the best choice.